AD A053192

AD No.
DDC FILE COPY

(12)

Massachusetts Institute of Technology
Research Laboratory of Electronics
Cambridge, Massachusetts 02139

Reprinted from

RLE Progress Report No. 120, January 1978

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

6 HOMOMORPHIC SPEECH ANALYSIS-SYNTHESIS.

5. TYPE OF REPORT & PERIOD COVERED

RLE Progress Report Reprint

6. PERFORMING ORG. REPORT NUMBER

7. ~~AUTHOR(s)~~

ALAN V./Oppenheim, Arthur B./Baggeroer
~~A. V. Oppenheim, et al.~~
James H./McClellan

8. CONTRACT OR GRANT NUMBER(s)

15 N00014-75-C-0951

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

NR 049-328

11. CONTROLLING OFFICE NAME AND ADDRESS

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

12. ~~REPORT DATE~~

11 Jan[redacted]78

13. NUMBER OF PAGES

12 4 5p.

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

Office of Naval Research
Information Systems Program
Code 437
Arlington, Virginia 22217

15. SECURITY CLASS. (of this report)

Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

9 Progress rept.

18. SUPPLEMENTARY NOTES

Reprinted from RLE Progress Report No. 120, Research Laboratory of Electronics, M.I.T., Cambridge, Mass., January 1978, Section XXI, pp. 103-105.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Digital signal processing
Vocoders
Homomorphic speech

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This Work by A. V. Oppenheim and his students and collaborators is summarized in the following projects: Homomorphic speech analysis-synthesis, enhancement of degraded speech, time-varying linear predictive coding of speech signals, and digital seismic signal processing.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

304 050

XXI. DIGITAL SIGNAL PROCESSING

Academic Research Staff

Prof. Alan V. Oppenheim
Prof. Arthur B. Baggeroer
Prof. James H. McClellan

Graduate Students

Bir Bhanu
Thomas E. Bordley
Patrick W. Bosshart
David S. K. Chan
Gregory L. Duckworth
Mark G. Hall
David B. Harris
Samuel Holtzman
Richard B. Kline
Stephen W. Lang
David C. LeDoux
Steven J. Leverette
Jae S. Lim
David R. Martinez
Thomas L. Marzetta
Syed H. Nawab
Robert W. Patterson
Michael R. Portnoff
Thomas F. Quatieri, Jr.
Antonio Ruiz
Kenneth B. Theriault
José M. Tribolet
Pirooz Vatan

1. HOMOMORPHIC SPEECH ANALYSIS-SYNTHESIS

U. S. Navy – Office of Naval Research (Contract N00014-75-C-0951)

Thomas F. Quatieri, Jr., Alan V. Oppenheim, Antonio Ruiz

a. Simulation of a Homomorphic Vocoder Based on Charged Coupled Device (CCD) Technology

We have completed simulations of various speech analysis-synthesis configurations based on both the conventional chirp-z-transform (CZT) realization of the discrete Fourier transform and the sliding CZT realization of the discrete sliding Fourier transform. These realizations are amenable to CCD technology and allow for real-time, low-cost implementation of the homomorphic vocoder.

A comparative study was performed, illustrating the tradeoffs between synthetic speech quality and implementational complexity for the two schemes.

b. Quality Improvement

Techniques for synthetic speech quality improvement were tested and evaluated in collaboration with studies of the Speech Group at Lincoln Laboratory. Issues such as interpolation, amplitude measurements, buzziness, hoarseness, and coding were explored. For male speech, formal listening tests indicate that for low bit rates (2400 to 3600 bps) the homomorphic system is comparable in quality to more established schemes such as LPC and the channel vocoder. For high bit rates (8000 to 9600 bps), the homomorphic system was judged to have the highest quality.

The female synthetic speech, on the other hand, unlike that of the male, tends in general to be degraded by a "hoarseness." We are investigating ways of rigorously characterizing this degradation and are exploring adaptive techniques for improvement.

We shall soon be completing a pitch-synchronous complex cepstral vocoder. This scheme, we hope, will yield an understanding of the effects of phase of the estimated vocal tract impulse response on synthetic speech quality.

2. ENHANCEMENT OF DEGRADED SPEECH

U. S. Navy – Office of Naval Research (Contract N00014-75-C-0951)

Jae S. Lim, Alan V. Oppenheim

Continuing our work on enhancing degraded speech, we have attempted to develop a complete analysis/synthesis system in which the synthesis parameters are estimated from noisy speech data. The particular analysis/synthesis system we have considered is based on an all-pole model of speech. Our approach has been to apply a Maximum A Posteriori (MAP) estimation procedure in estimating the coefficient vector of an all-pole system from noisy speech accounting for the presence of noise. In general, a MAP estimation procedure for noisy speech leads to solving a set of nonlinear equations. Two suboptimal procedures which require solving only sets of linear equations have, however, been developed. These methods have been applied to both synthetic and real speech data with white Gaussian background noise, and our preliminary listening test indicates that both systems are capable of significant noise reduction. We are now engaged in a more formal subjective test which is directed toward evaluating the two linear systems in terms of their performance in enhancing speech intelligibility and quality when the background noise is of various different spectra.

3. TIME-VARYING LINEAR PREDICTIVE CODING OF SPEECH SIGNALS

U. S. Navy – Office of Naval Research (Contract N00014-75-C-0951)

Mark G. Hall, Alan V. Oppenheim, Alan S. Willsky
[Prof. Willsky is Assistant Director of the Electronic Systems Laboratory, M. I. T.]

During this year we have completed a project involved with time-varying predictive coding of speech. The project involved a generalization of linear prediction using time-varying coefficients. By representing each time-varying coefficient, either in terms of a power series, or in terms of a Fourier series, a set of equations to determine the coefficients was obtained. These coefficients are reminiscent of those in the time-invariant case in that they are block-symmetric or block-Toeplitz. The basic problem can either be formulated in a covariance form or in a correlation form and the relative characteristics of these two approaches were explored. Through a study of a number of synthetic examples and examples using real data, we concluded that the

covariance form with a power series representation for the coefficients was the most preferable and that this approach has the potential for representing a long nonstationary segment of speech with fewer total coefficients than would be required through the use of time-invariant LPC in which an analysis window is moved through the data.

4. DIGITAL SEISMIC SIGNAL PROCESSING

U. S. Navy – Office of Naval Research (Contract N00014-75-C-0951)
National Science Foundation (Grant ENG76-24117)

David B. Harris

A seismic surveying technique called wave equation migration is being investigated for possible applications of two-dimensional digital signal processing algorithms. The key component of the migration algorithm is a difference equation approximation to the wave equation. This difference equation is used to extrapolate a wave field recorded on the boundary of a region backwards into the region. An ideal transfer function for the two-dimensional difference equation can be derived from the wave equation. Currently, methods are being sought to approximate this transfer function which is all-pass with a specified phase.

5. TIME SCALE MODIFICATIONS OF SPEECH SIGNALS

Michael R. Portnoff

The objective of our research in this area is to modify a speech signal in such a manner that the resulting signal is perceived as identical to the original except for its rate of articulation. In particular, we seek to preserve such qualities as naturalness, intelligibility, and speaker-dependent features, while avoiding the introduction of such objectionable artifacts as "glitches," "burbles," and reverberation often present in vocoded speech.

We have developed and demonstrated a high-quality system for time-scale compression and expansion of speech based on short-time Fourier analysis.[1, 2] This system is capable of compressing speech by ratios as large as 3:1 and expanding speech by arbitrarily large ratios. Furthermore, the performance of this system does not appear to be sensitive to the presence of broadband noise in the speech source material.

References

1. M. R. Portnoff, "Implementation of the Digital Phase Vocoder Using the Fast Fourier Transform," IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-24, No. 3, pp. 245-248, June 1976.

2. M. R. Portnoff, "A Mathematical Framework for Time-Scale Modification of Speech Signals," Ninety-third Meeting, Acoustical Society of America, State College, Pennsylvania, June 7-10, 1977. (Abstract in J. Acoust. Soc. Am., Vol. 61, Suppl. No. 1, Spring 1977, p. S68.)

6. ALGORITHMS FOR HIGHLY PARALLEL COMPUTER STRUCTURES

National Aeronautics and Space Administration (Grant NSG-5157)

James H. McClellan, David C. LeDoux

In order to process photographic images from Earth observation satellites rapidly, computers with a very high degree of parallel processing capability are being designed. One such machine[1] would be composed of a minimum of 16, 384 very simple processors arranged in a 128 × 128 array. Each processing element contains a bit-serial adder with a small amount of memory and is able to transfer data, one bit at a time, to any of its four nearest neighbors. A central control unit issues an instruction to the entire array and the individual processors decide whether or not to execute it, depending on a mask bit in each processor. In use, each processor operates on one pixel in a sampled photo and the parallel processing allows great time savings over more traditional serial computers.

A frequent operation in satellite image processing is that of image registration which involves computing the cross-correlation between two images of the same scene. Correlations may be computed indirectly and sometimes more efficiently by using two-dimensional transforms such as the Fourier transform or the Fermat number transform (FNT). The computation of the FNT does not involve multiplications or complex numbers, and is well suited to the simplicity of the small processors. We have investigated the use of the FNT to perform correlation on a highly parallel computer developed by NASA. Programs have been writen and timing estimates have been obtained. The FNT can be more efficient than direct computation of the correlation (on this machine), depending on the sizes of the images used.

We are now investigating the applicability of the parallel computer structure to several new image restoration algorithms, particularly those which attempt to correct for the effects of the Earth's turbulent atmosphere.

References

1. L. W. Fung, MPPC: A Massively Parallel Processing Computer, NASA Goddard Space Flight Center, September 1976.

Distribution List

Contract N00014-75-C-0951

	No. copies
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	12
Office of Naval Research Information Systems Program Code 437 Arlington, Virginia 22217	2
Office of Naval Research Code 105 Arlington, Virginia 22217	6
Office of Naval Research Branch Office/Boston 495 Summer Street Boston, Massachusetts 02210	1
Office of Naval Research Branch Office/Chicago 536 South Clark Street Chicago, Illinois 60605	1
Office of Naval Research Branch Office/Pasadena 1030 East Green Street Pasadena, California 91106	1
Naval Research Laboratory Technical Information Division, Code 2627 Washington, D.C. 20375	6
Commandant of the Marine Corps (Code AX) Dr. A. L. Slafkosky Scientific Advisor Washington, D.C. 20380	1
Office of Naval Research Code 455 Arlington, Virginia 22217	1
Office of Naval Research Code 458 Arlington, Virginia 22217	1
Naval Electronics Laboratory Center Advanced Software Technology Division Code 5200 San Diego, California 92152	1

Mr. G. H. Gleissner 1
Naval Ship Research and Development Center
Computation and Mathematics Department
Bethesda, Maryland 20034

~~Dr. J. Laderman~~ 1
~~Office of Naval Research~~
~~New York Area Office~~
~~207 West 24th Street~~
~~New York, New York 10011~~

Captain Grace M. Hopper 1
Office of Chief of Naval Operations
NAICOM/MIS Planning Branch
NOP-916D Pentagon
Washington, D.C. 20350

Advanced Research Projects Agency 1
Information Processing Techniques
1400 Wilson Boulevard
Arlington, Virginia 22209

Commander 1
U.S. Army Electronics Command
Attn: Mr. J. L. DeClerk
AMSEL-NLY
Fort Monmouth, New Jersey 07703

Dr. Robert Kahn 1
Program Manager
Information Processing Techniques
1400 Wilson Boulevard
Arlington, Virginia 22209

Mr. James M. White 1
IBM
Thomas J. Watson Research Center
P. O. Box 218
Yorktown Heights, New York 10598

ONR New York Area Office **1**
715 Broadway, 5th Floor
New York New York 10003